

LINEAIRE STRUCTURELE MODELLEN MET RUIMTELIJK-TEMPORELE CORRELATIE

Henk Folmer
Geografisch Instituut Rijksuniversiteit Groningen
Postbus 800, 9700 AV Groningen

Peter Nijkamp
Economische Faculteit, Vrije Universiteit Amsterdam
Postbus 7161, 1007 MC Amsterdam.

SAMENVATTING

In regionaal-economisch en geografisch onderzoek worden in toenemende mate steekproeven geanalyseerd op basis van waarnemingen uit R regio's over T perioden. Bij dergelijke steekproeven dient men rekening te houden met 3 vormen van correlatie:

- ruimtelijke autocorrelatie;
- ruimtelijke crosscorrelatie;
- temporele autocorrelatie.

Bij het ruimtelijk onderzoek bestaat ook een groeiende belangstelling voor het gebruik van lineaire structurele modellen.

In dit artikel wordt aangegeven op welke wijze rekening kan worden gehouden met ruimtelijke en temporele correlatie bij genoemde typen modellen. Bovendien wordt aandacht besteed aan de opsporing van ruimtelijke cross-correlatie waarvoor geen toetsen beschikbaar zijn.

1. INLEIDING

Lineaire structurele modellen (afgekort als LISREL modellen) bieden de mogelijkheid om expliciet rekening te houden met latente variabelen ¹⁾ en genieten vooral om die reden een toenemende belangstelling in de ruimtelijke economie en geografie. Dit type modellen is toegepast onder andere in

- 1) Latente variabelen, ook wel theoretische constructen genoemd, vormen de groep van wetenschappelijke begrippen die betrekking hebben op verschijnselen die niet direkt waarneembaar zijn. Voorbeelden zijn: 'welvaart', 'permanent inkomen', 'economische verwachting', etc. Latente variabelen vormen de bouwstenen van wetenschappelijke theorieën en derhalve dienen zij opgenomen te worden in een empirische analyse (zie voor verdere informatie, o.a., McCorguodale en Weehl, 1956, en Hempel, 1958). Bovendien kan het gebruik van latente variabelen het multicollineariteitsprobleem aanzienlijk reduceren.

het kader van theorievorming (Droth en Fischer, 1981) en van de meting van effecten van regionaal beleid (Folmer, 1981, en Folmer en Oosterhaven, 1982).

Gewoonlijk gaat men bij het gebruik van LISREL modellen uit van onafhankelijke waarnemingen. In geval van toepassingen in genoemde vakgebieden dient men echter rekening te houden met de mogelijkheid van ruimtelijk-temporele correlatie. In dit artikel wordt een methode besproken om deze vorm van correlatie te verdisconteren bij het modelleren van ruimtelijke verschijnselen. Daartoe worden in sectie 2 de structuur, identificatie, schatting en evaluatie van LISREL modellen in het algemeen samengevat. In sectie 3 komen de aard en de identificatie van ruimtelijke auto- en crosscorrelatie aan de orde. Tenslotte, de voorgestelde methode om ruimtelijk-temporele correlatie te verdisconteren in LISREL modellen wordt behandeld in sectie 4.

2. DE BELANGRIJKSTE KENMERKEN VAN LISREL MODELLEN

De klasse van LISREL modellen is ontwikkeld door onder andere Jöreskog (1973a, 1973b, 1977) en Goldberger en Duncan (1973). Schatten en evalueren van deze modellen kan gedaan worden met het LISREL V computerprogramma (Jöreskog en Sörbom, 1981). Deze sectie is grotendeels gebaseerd op deze referenties.

Alvorens het model te beschrijven merken we op, dat het mogelijk is constanten op te nemen in de hieronder te presenteren modellen, hetgeen vooral van belang is bij de vergelijking van verschillende verzamelingen regio's. In geval van de analyse van één steekproef verschaffen de constanten meestal weinig informatie. Gemakshalve laten we de constanten hier achterwege en veronderstellen we, dat zowel de observeerbare als latente variabelen gecentreerd zijn.

We definiëren $y = (y_1, y_2, \dots, y_p)^T$ en $x = (x_1, x_2, \dots, x_q)^T$ als vectoren van waarneembare variabelen, $\eta = (\eta_1, \eta_2, \dots, \eta_m)^T$ en $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ als vectoren van latente variabelen, en $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_p)^T$ en $\delta = (\delta_1, \delta_2, \dots, \delta_q)^T$ als vectoren van meetfouten van respectievelijk y en x .

Het verband tussen de observeerbare en latente variabelen is gegeven in de meetmodellen (1) en (2) :

$$y = \Lambda_y \eta + \epsilon \quad (1)$$

en

$$x = \Lambda_x \xi + \delta \quad (2)$$

waarbij Λ_y ($p \times m$) en Λ_x ($q \times n$) matrices van regressie coëfficiënten zijn.

Het structurele model bestaat uit relaties tussen de latente variabelen:

$$\eta = \tilde{B} \eta + \Gamma \xi + \zeta \quad (3)$$

oftewel:

$$B \eta = \Gamma \xi + \zeta \quad (4)$$

waarbij $\tilde{B}$ ($m \times m$) en Γ ($m \times n$) matrices van structurele coëfficiënten zijn en ζ ($m \times 1$) een vector van storings is. Bovendien is $B = I - \tilde{B}$, waarbij I de eenheidsmatrix is.

De volgende veronderstellingen worden gemaakt:

$$\left. \begin{aligned} E(y) &= 0 ; E(x) = 0 ; E(\eta) = 0 ; E(\xi) = 0 ; E(\epsilon) = 0 & E(\delta) = 0 \\ E(\zeta) &= 0 \\ E(\eta \epsilon^T) &= 0 , E(\xi \delta^T) = 0 , E(\eta \delta^T) = 0 , E(\xi \epsilon^T) = 0 , E(\epsilon \delta^T) = 0 \\ E(\zeta \epsilon^T) &= 0 , E(\zeta \delta^T) = 0 , E(\zeta \epsilon^T) = 0 \end{aligned} \right\} \quad (5)$$

De covariantiematrices van ϵ en δ , welke bij LISREL modellen niet diagonaal behoeven te zijn, worden hier aangeduid als θ_ϵ ($p \times p$) en θ_δ ($q \times q$) en de covariantiematrices van ξ en ζ als Φ ($n \times n$) en Ψ ($m \times m$).

We maken hier de volgende opmerkingen. In de eerste plaats, meerdere observeerbare variabelen voor één latente variabele zijn vaak noodzakelijk in verband met de identificatie. Bovendien kan één observeerbare variabele een indicator zijn van meerdere latente variabelen. Op de tweede plaats, door bepaalde specificaties kunnen diverse submodellen van model (1) - (4) verkregen worden. We noemen hier in het bijzonder de fixed-x optie waarbij de conditionele verdeling van de y variabelen voor gegeven x bestudeerd wordt. In dat geval wordt gespecificeerd: $\xi \equiv x$, $\Lambda_x = I$, $\theta_\delta = 0$. Dit type model wordt veel gebruikt in econometrische toepassingen.

Laten we nu overgaan tot de behandeling van de identificatie-, schattings- en toetsingsproblemen van model (1) - (4). Daartoe besteden we eerst aandacht aan de theoretische covariantiematrix $\Sigma ((p \times q) \times (p \times q))$ van $z = (y^T, x^T)^T$. Met de boven beschreven notaties en veronderstellingen kan deze matrix geschreven worden als :

$$\Sigma = \begin{bmatrix} \Lambda_y (B^{-1} \Gamma \Phi \Gamma^T (B^{-1})^T + B^{-1} \Psi (B^{-1})^T) \Lambda_y^T + \theta_\epsilon & \Lambda_y B^{-1} \Gamma \Phi \Lambda_x^T \\ \Lambda_x \Phi \Gamma^T (B^{-1})^T \Lambda_y^T & \Lambda_x \Phi \Lambda_x^T + \theta_\delta \end{bmatrix} \quad (6)$$

Op basis van theoretische overwegingen, verwachtingen, etc. worden bepaalde elementen in de parameter matrices, en dus in Σ , beschouwd als vrije, gefixeerde, of beperkte parameters. Het laatste betekent, dat de desbetreffende parameters onbekend zijn, doch als gelijk aan een of meerdere andere parameters beschouwd worden. Alle onafhankelijke, vrije en beperkte parameters in de matrices Λ_x , Λ_y , B , Γ , Φ , Ψ , θ_ϵ en θ_δ zullen we aangeven met θ en de corresponderende Σ met $\Sigma(\theta)$.

Laten we nu aandacht besteden aan de steekproef covariantiematrix S van z . Als de observeerbare variabelen gemeten zijn op een interval of ratio schaal is deze gedefiniëerd als :

$$S = \frac{1}{M-1} \sum_{i=1}^M (z_i - \bar{z})(z_i - \bar{z})^T \quad (7)$$

waarbij M het aantal waarnemingen voorstelt en $\bar{z}$ de vector van steekproefgemiddelden.

We merken op, dat het LISREL V programma 3 soorten steekproef matrices kan analyseren: de produktmomenten-, de correlatie-, en de covariantiematrix. (Wanneer welk type matrix gebruikt moet worden, is beschreven door Jöreskog en Sörbom, 1981.) In dit verband willen we er op wijzen, dat de covariantiematrix van de schatters van de onbekende parameters gebaseerd is op de analyse van een steekproef-covariantiematrix. Dit betekent, dat indien een ander type steekproefmatrix geanalyseerd wordt, de betrouwbaarheidsintervallen voor de diverse parameters met grote voorzichtigheid geïnterpreteerd moeten worden.

Ook wanneer de geobserveerde variabelen niet van interval- of ratio-niveau zijn, kan het LISREL V programma in een veelheid van gevallen gebruikt worden. In geval van nominale of ordinale x-variabelen kan de fixed-x optie gebruikt worden. Wanneer er ordinale variabelen voorkomen tussen de y-variabelen, of tussen de x-variabelen, die niet als 'fixed' beschouwd dienen te worden, kan het LISREL V programma polychorische of polyseriale correlatie coëfficiënten analyseren. In geval van nominale y, of x vari-

abelen, die niet als fixed behandeld mogen worden, kan de steekproef in meerdere groepen verdeeld worden, welke vervolgens simultaan geanalyseerd kunnen worden.

Schatten van LISREL modellen geschiedt door het minimaliseren van de afstand tussen S en $\Sigma(\theta)$ in een bepaalde metriek met betrekking tot θ . Bij de schattingsprocedure wordt de veronderstelling gemaakt dat de verdeling van de observeerbare variabelen beschreven kan worden met behulp van de momenten van de eerste en tweede orde. Omdat aangenomen is dat de variabelen gecentraliseerd zijn, leidt deze veronderstelling ertoe dat de verdeling van z gekarakteriseerd is door de onafhankelijke parameters in $\Sigma(\theta)$. Vandaar dat schatten van LISREL modellen gebeurt door het minimaliseren van de afstand tussen S en $\Sigma(\theta)$.

Teneinde de onbekende parameters te kunnen schatten moet het model geïdentificeerd zijn, hetgeen wil zeggen dat de vector θ uniek bepaald moet zijn door $\Sigma(\theta)$. Een eerste noodzakelijke voorwaarde voor identificatie is dat het aantal verschillende elementen in $\Sigma(\theta)$ minstens even groot is als het aantal te schatten parameters. Een tweede noodzakelijke voorwaarde is dat iedere individuele parameter gescheiden kan worden van de overige parameters. Deze voorwaarde, die evenmin als de eerste voldoende is, is moeilijk te onderzoeken. Echter, het LISREL programma geeft aanwijzingen over identificatieproblemen. Het berekent een schatting van de matrix van tweede orde afgeleiden met betrekking tot de onbekende parameters van de fitting functie welke gebruikt wordt om het model te schatten. Als deze matrix singulier is, is het model meestal niet geïdentificeerd (cf. Silvey, 1970). Het LISREL programma geeft ook aan welke parameters niet geïdentificeerd zijn. In geval van modellen met latente variabelen is het model niet geïdentificeerd, als er geen meetschalen zijn vastgesteld voor de latente variabelen. Dit probleem kan eenvoudig opgelost worden door voor iedere latente variabele één λ -coëfficiënt gelijk te stellen aan 1.

Laten we nu aandacht besteden aan het eigenlijke schatten. Het LISREL V programma biedt 3 consistente schatters:

- initiële schatters;
- maximum likelihood (ML);
- ongewogen kleinste kwadraten (ULS).

Het eerste type levert startwaarden voor de iteratieve ML en ULS procedures. Omdat de initiële schattingsprocedure relatief weinig rekentijd vergt en de schattingen meestal weinig afwijken van de resultaten van de beide overige

schatteurs, kan men ook met de initiële schattingen volstaan.

De ML procedure geeft 'echte' maximum likelihood schatters als de geobserveerde variabelen multinormaal verdeeld zijn. Bovendien geeft deze methode voor een grote klasse van niet-normaal verdeelde variabelen goede schatters (Dijkstra, 1981).

Tenslotte, de ULS methode kan gebruikt worden zonder veronderstellingen ten aanzien van de verdeling van de variabelen. In tegenstelling tot ML behoeft bij de ULS procedure noch S , noch $\Sigma(\hat{\theta})$, waarbij $\hat{\theta}$ de startwaarden voor het optimaliseringsalgoritme zijn, positief definit te zijn.

Bij de bespreking van ruimtelijke crosscorrelatie zullen de geschatte residuen ter sprake komen. Deze definiëren we als :

$$e_m = y'_m - \hat{y}_m \quad (8)$$

waarbij y'_m en $\hat{y}_m$ de vectoren van waarden, respectievelijk van schattingen van de y -variabelen bij de m^e waarneming zijn. De residuen worden niet gegeven door het LISREL V programma, maar kunnen berekend worden met behulp van :

$$\hat{y}_m = \hat{\Lambda}_y \hat{B}^{-1} \hat{\Gamma} \hat{\Phi} \hat{\Lambda}_x (\hat{\Lambda}_x \hat{\Phi} \hat{\Lambda}_x^T + \hat{\theta}_{\delta})^{-1} x_m \quad (9)$$

Zie voor een afleiding hiervan Folmer en Nijkamp, 1982.

Ter afsluiting van deze paragraaf besteden we aandacht aan de evaluatie van een geschat model. Hierbij maken we onderscheid tussen het 'echte' toetsen van hypothesen volgens de regels van de statistische beslissingstheorie en het beoordelen van de fit van een model. In het eerste geval is een bepaalde hypothese beschikbaar welke aan de hand van de steekproefgegevens getoetst wordt. In het tweede geval wordt gezocht naar een model dat zo goed mogelijk past bij de beschikbare data. Daartoe worden modellen met verschillende specificaties geschat op basis van hetzelfde steekproefmateriaal. In deze situatie is model evaluatie in de vorm van hypothesetoetsen minder geschikt, omdat de nauwkeurigheid van de schatter van een dergelijk op de data geïnspireerd model overschat wordt (zie, bijvoorbeeld Leamer, 1978, en Dijkstra, 1981).

De steekproefgrootheden welke door het LISREL programma gegeven worden en welke gebruikt kunnen worden voor beoordeling van de fit en/of voor toetsing van hypothesen hebben betrekking op :

- individuele parameters;
- de afzonderlijke vergelijkingen van de meetmodellen;
- de afzonderlijke vergelijkingen van het structurele model (zie Folmer en Nijkamp, 1982, voor een kritiek);
- het meetmodel voor de endogene en exogene latente variabelen gezamenlijk;
- het structurele model (zie Folmer en Nijkamp, 1982, voor een kritiek).
- het model als geheel.

Een gebrekkige fit van een model kan het gevolg zijn van :

- parameters welke bijna niet geïdentificeerd zijn;
- specificatiefouten met betrekking tot de verdeling der geobserveerde variabelen;
- parameters welke ten onrechte gefixeerd zijn (gewoonlijk gelijk aan nul);
- parameters welke ten onrechte gespecificeerd zijn als vrije parameters;
- specificatiefouten met betrekking tot de vorm van het model;
- ontbrekende variabelen.

Als vuistregel kan men aanhouden, dat een wezenlijke verbetering van een model weerspiegeld wordt in een wezenlijke verbetering van de bovengenoemde evaluatie steekproefgrootheden (voor een overzicht van indicaties om de diverse soorten specificatiefouten te herkennen en te herstellen, zie Folmer en Nijkamp, 1982).

De standaard steekproefgrootheden van het LISREL programma, zoals de likelihood-ratio toetsgrootheid en de standaardfouten van de individuele onbekende parameters kunnen op de gebruikelijke wijze voor 'echte' toetsingsdoel-einden gebruikt worden als aan de volgende voorwaarden is voldaan:

- de latente variabelen en meetfouten zijn multi-normaal verdeeld;
- de steekproefomvang is voldoende groot;
- een covariantiematrix is geanalyseerd;
- de ML schattingsprocedure is gebruikt.

Op de eerste plaats kan met behulp van een likelihood-ratio toets getoetst worden de hypothese:

$H_0 : \Sigma = \Sigma(\theta)$, waarbij θ de structuur van het onderhavige model is,
versus

$H_a : \Sigma$ is niet beperkt.

Op de tweede plaats, een rij van 'geneste' hypothesen kan ook getoetst worden met een likelihood-ratio toets.

Tenslotte, er kunnen betrouwbaarheidsintervallen voor de individuele schatters geconstrueerd worden. Onder de genoemde condities is de gestandaardiseerde schatter van iedere parameter asymptotisch standaard normaal verdeeld.

3. AARD EN IDENTIFICATIE VAN RUIMTELIJK-TEMPORELE AUTO- EN CROSS-CORRELATIE

Veronderstel dat een steekproef bestaande uit R regio's in T periodes beschikbaar is. Bij een dergelijke steekproef kunnen zich 3 vormen van correlatie tussen de storingen en meetfouten voordoen:

- temporele autocorrelatie;
- ruimtelijke autocorrelatie;
- ruimtelijke crosscorrelatie.

Temporele autocorrelatie is uitvoerig beschreven in de literatuur (zie onder andere INSEE, 1978); vandaar dat we ons hier beperken tot een bespreking van de andere twee vormen van correlatie.

In geval van ruimtelijke autocorrelatie wordt een verschijnsel in een bepaalde regio, bijv. regio r , beïnvloed door hetzelfde verschijnsel in andere regio's in de lopende of in voorgaande periodes. Bovendien kan er een omgekeerde invloed bestaan: het verschijnsel in andere regio's kan beïnvloed worden door het verschijnsel in regio r in de lopende of in voorgaande periodes. In geval van 2 verschillende verschijnselen spreken we van ruimtelijke crosscorrelatie.

Ruimtelijke auto- en crosscorrelatie kan onder meer gemeten worden met behulp van de Moran coëfficiënt. Hierbij wordt gespecificeerd naar de mate van aangrenzendheid oftewel contiguiteitsorde (bijv. s) en de temporele vertraging (bijv. l).

In geval van autocorrelatie is de betreffende coëfficiënt gedefinieerd als:

$$M_{l,1}^s(y,y) = \frac{\sum_{r=1}^R (y_{r,t} - \bar{y}_t) (L^s y_{r,t-l} - \bar{y}_{t-l})}{\left\{ \sum_r (y_{r,t} - \bar{y}_t)^2 \right\}^{\frac{1}{2}} \left\{ \sum_r (y_{r,t-l} - \bar{y}_{t-l})^2 \right\}^{\frac{1}{2}}} \quad (10)$$

$$s = 1, 2, \dots, S$$

$$L = 0, 1, 2, \dots, T$$

met

$y_{r,t}$ de variabele in studie in regio r op tijdstip t ;

L^s de ruimtelijke vertragingoperator zodanig dat:

$$L^s y_{r,t} = \sum_{i \in A_{s,r}} w_{r,i}^s y_{i,t}, \quad (11)$$

waarbij $A_{s,r}$ de verzameling is van alle regio's van contiguiteitsorde s met betrekking tot regio r en $w_{r,i}^s$ een contiguiteitsgewicht tussen de regio's r en i zodanig dat:

$$\sum_{i \in A_{s,r}} w_{r,i}^s = 1, \quad \forall r, s \quad (12)$$

$$\bar{y}_t = \frac{1}{R} \sum_{r=1}^R y_{r,t}$$

De Moran coëfficiënt voor ruimtelijke crosscorrelatie voor twee variabelen x en y voor contiguiteitsorde s en temporele vertraging l is gedefiniëerd als

$$M_1^s(y, x) = \frac{\sum_{r=1}^R (y_{r,t} - \bar{y}_t) (L^s x_{r,t-l} - \bar{x}_{t-l})}{\left\{ \sum_r (y_{r,t} - \bar{y}_t)^2 \right\}^{\frac{1}{2}} \left\{ \sum_r (x_{r,t-l} - \bar{x}_{t-l})^2 \right\}^{\frac{1}{2}}} \quad (13)$$

$$\begin{aligned} s &= 1, 2, \dots, S \\ l &= 0, 1, 2, \dots, T \end{aligned}$$

(voor nadere details verwijzen we naar Cliff en Ord, 1973, Martin en Oeppen, 1975, en Hordijk en Nijkamp, 1977).

Voor iedere variabele en voor ieder paar variabelen kan voor iedere temporele vertraging en voor iedere contiguiteitsorde een Moran coëfficiënt berekend worden. Aldus verkrijgt men een matrix van ruimtelijke auto- en crosscorrelaties (zie Martin en Oeppen, 1975, en Hordijk en Nijkamp, 1977).

Over het algemeen zijn de momenten van Moran coëfficiënten complexe functies. Cliff en Ord (1973) en Haggett et al. (1973) hebben de momenten van de eerste en tweede orde afgeleid van $M_1^s(x, y)$ als de betreffende variabele normaal verdeeld is onder de veronderstelling van afwezigheid van ruimtelijke autocorrelatie. Bovendien hebben zij aangetoond, dat de betreffende coëfficiënt normaal verdeeld is onder de genoemde condities. Dit betekent, dat de hypothese van afwezigheid van ruimtelijke autocorrelatie op de gebruikelijke wijze kan worden getoetst.

Alvorens de identificatie van ruimtelijke cross-correlatie te bespreken merken we het volgende op. Ten eerste, deze vorm van correlatie kan veroorzaakt zijn door de verklarende variabelen welke expliciet zijn opgenomen in het model en/of door variabelen gerepresenteerd door de storingsen. Ten tweede, de momenten van $M_1^s(x,y)$ zijn niet bekend, zodat deze coëfficiënt niet voor toetsingsdoeleinden kan worden gebruikt. Dit betekent, dat ruimtelijke cross-correlatie op een indirecte wijze opgespoord moet worden. We beginnen met ruimtelijke cross-correlatie tussen de y-variabelen onderling en tussen de y-en x-variabelen. Het aantal van de y-en x-variabelen nemen we gelijk aan n. We stellen nu de volgende procedure voor:

- schat het model zonder specificaties voor ruimtelijke cross-correlatie, maar met specificaties voor eventuele ruimtelijke en temporele autocorrelatie op de wijze zoals beschreven in sectie 4;
- bereken de residuen met behulp van (8) en (9);
- toets de residuen op ruimtelijke auto-correlatie met behulp van (10);
- als de hypothese van ruimtelijk ongecorreleerde residuen niet verworpen wordt, behoeft ruimtelijke cross-correlatie niet verder beschouwd te worden;
- als de hypothese van ruimtelijk ongecorreleerde residuen verworpen wordt, bereken dan de matrix C van ruimtelijke cross-correlatie-coëfficiënten volgens (13);
- het cross-correlatie-effect van de variabele corresponderend met het element van deze matrix, dat het grootst is in absolute waarde, wordt behandeld op de wijze zoals beschreven in sectie 4;
- het gewijzigde model wordt herschat en als de coëfficiënt van de variabele, welke het ruimtelijke cross-correlatie-effect representeert, significant verschilt van nul, wordt deze variabele in het model opgenomen; het aantal ruimtelijk gecorreleerde variabelen daalt hierdoor met 1;
- anders wordt het in absolute waarde op één na grootste element van de matrix C van cross-correlatie-coëfficiënten en de corresponderende x variabele beschouwd; het aantal eventueel ruimtelijk gecorreleerde variabelen wordt hierdoor verminderd met 1.
- deze zoekprocedure wordt voortgezet, totdat een significant van nul verschillende coëfficiënt van een variabele die ruimtelijke cross-correlatie representeert gevonden wordt en zolang het aantal expliciet in het model opgenomen variabelen dat ruimtelijk gecorreleerd kan zijn, niet uitgeput is.
- wanneer een significant van nul verschillende coëfficiënt gevonden is, worden de residuen van het model, uitgebreid met de cross-correlatie variabele, geschat, getoetst op ruimtelijke autocorrelatie, enzovoort.

De bovenstaande procedure is samengevat in figuur 1. Hierbij zijn ruimtelijke auto- en cross-correlatie afgekort als $r a c$, respectievelijk $r c c$. Bovendien is $\tilde{C}$ gedefinieerd als de matrix met als i, j^e element c_{ij} van de matrix C en nullen elders.

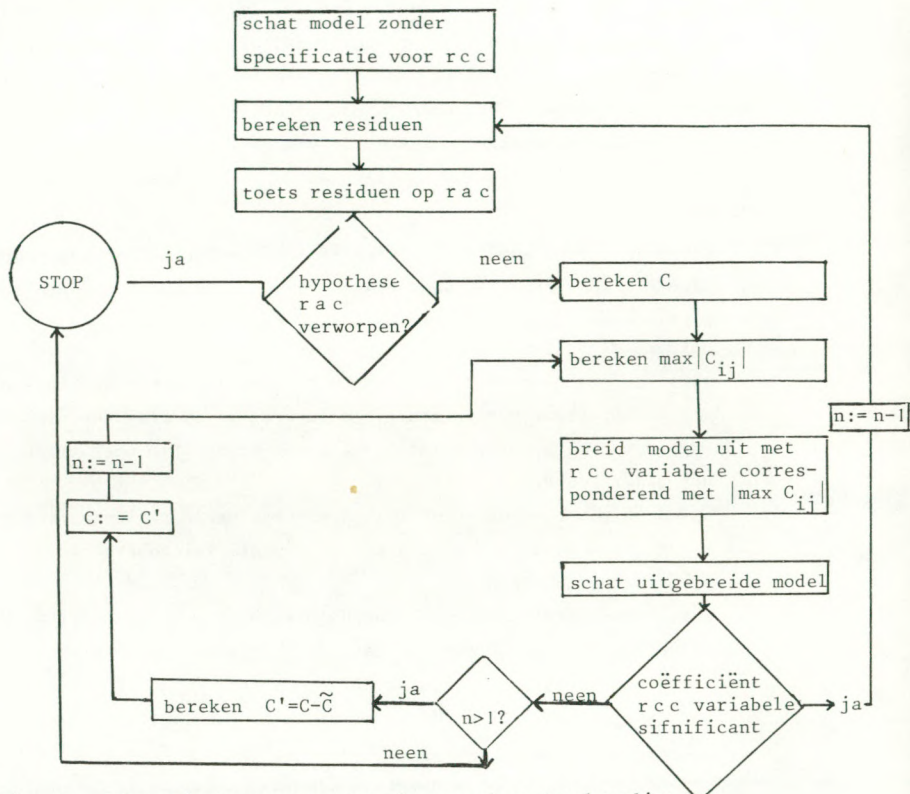


Fig. 1. Schematische weergave van de procedure ter bepaling van ruimtelijke cross-correlatie.

Wanneer de boven beschreven procedure stopt omdat het aantal relevante variabelen uitgeput is en de residuen nog ruimtelijk gecorreleerd zijn, moet er sprake zijn van ruimtelijke auto- of cross-correlatie in de variabelen gepresenteerd door de storingsen. In de volgende paragraaf zal worden aangegeven hoe met deze vorm van ruimtelijke correlatie rekening kan worden gehouden.

Tot slot van deze paragraaf beschouwen we een algemene specificatie van een model met ruimtelijke auto- en cross-correlatie.

$$\begin{aligned}
 y_{r,t} = & \sum_{j=1}^T \alpha_j^r y_{r,t-1} + \sum_{s=1}^S \sum_{l=0}^L \beta_l^s (L^s y_{r,t-1}) + \sum_{j=1}^J \sum_{l=0}^{T_j'} Y_{j,l}^r x_{r,j,t-1} \\
 & + \sum_{j=1}^J \sum_{p=1}^P \sum_{l=0}^{\tilde{T}_j} \delta_{j,l}^p (L^p x_{r,j,t-1}) + \zeta_{r,t}
 \end{aligned} \quad (14)$$

met :

$y_{r,t}$ de endogene variabele in regio r op tijdstip t ;

$x_{r,j,t}$ de j -de exogene variabele in regio r op tijdstip t ;

$\alpha, \beta, \gamma, \delta$ onbekende parameters ;

$\zeta_{r,t}$ de verstoring.

De index 1 heeft betrekking op temporele vertragingen.

Schatten van een model van het type (14) kan belemmerd worden door multicollineariteit. Bovendien is het aantal waarnemingen in regionaal onderzoek vaak tamelijk klein, hetgeen een nadelige invloed kan hebben op het aantal vrijheidsgraden. In de volgende paragraaf zullen we laten zien, dat de LISREL benadering met betrekking tot deze problemen aantrekkelijke eigenschappen bezit.

4. SCHATTEN VAN RUIMTELIJK-TEMPORELE LISREL MODELLEN

We nemen weer aan dat we over een steekproef bestaande uit R regio's in T perioden beschikken en beschouwen het volgende LISREL model:

$$\left. \begin{aligned}
 B \eta_{r,t} &= \Gamma^1 \xi_{r,t} + \Gamma^2_t \tilde{\xi}_r + u_{r,t} & (a) \\
 y_{r,t} &= \Lambda_y \eta_{r,t} + \epsilon_{r,t} & (b) \\
 x_{r,t} &= \Lambda_x \xi_{r,t} + \delta_{r,t} & (c) \\
 z_r &= \Lambda_z \tilde{\xi}_r + \tilde{\delta}_r & (d)
 \end{aligned} \right\} \begin{array}{l} r=1,2,\dots,R \\ t=1,2,\dots,T \end{array} \quad (15)$$

met :

$\xi_{r,t}$ een vector van verklarende latente variabelen welke zowel lopende of vertraagde exogene, als vertraagde endogene variabelen kunnen zijn ;

$\tilde{\xi}_r$ een vector van verklarende variabelen welke in de tijd constant zijn;

z_r een vector van indicatoren voor $\tilde{\xi}_r$;

$u_{r,t}$ een vector van verstoringen;

Alle andere variabelen en matrices zijn gedefiniëerd als in paragraaf 2.

We besteden nu eerst aandacht aan de representatie van ruimtelijke auto- en cross-correlatie. Veronderstel, dat de j^e exogene variabele x_j ruimtelijk gecorreleerd is met y_i voor verschillende ruimtelijke en temporele vertragingen. Deze vorm van ruimtelijke correlatie nemen we op in het model via een nieuwe exogene latente variabele, zeg $\xi_{n',r,t}$ waarvan $L^P x_{r,t-1}$ indicatoren zijn voor de verschillende relevante waarden van p en l . Dit betekent dat de verzameling variabelen $\{L^P x_{r,t-1}, p=1,2,\dots,P; l=0,1,2,\dots,L\}$ in model (14) vervangen wordt door één latente variabele, hetgeen tot een aanzienlijke reductie van multicollineariteit kan leiden.

We merken in dit verband het volgende op. Ten eerste, wanneer men over aanwijzingen beschikt, dat de reeks $\{L^P x_{r,t-1}, p=1,2,\dots,P; l=0,1,2,\dots,L\}$ ruimtelijke of temporele 'breuken' vertoont, zodat er sprake is van verschillende deelreeksen, kan men meerdere latente variabelen gebruiken, corresponderend met ieder der deelreeksen, om ruimtelijke correlatie te representeren.

Ten tweede, als meerdere x variabelen voor diverse ruimtelijke en temporele vertragingen gecorreleerd zijn met y_i kunnen de betreffende x variabelen veelal gebruikt worden als indicatoren van één en dezelfde latente ruimtelijke correlatie variabele.

Een soortgelijke benadering kan gebruikt worden als er sprake is van ruimtelijke auto- of cross-correlatie tussen de endogene y variabelen. In dat geval worden één of meerdere latente endogene variabelen ingevoerd om de betreffende ruimtelijke correlatie te representeren. Omdat het niet de bedoeling is dergelijke endogene variabelen te verklaren in termen van de andere variabelen in het model, worden deze zogenaamde quasi-endogene variabelen in het structurele model gelijk gesteld aan de daarmee corresponderende storing.

De hier beschreven procedures kunnen toegepast worden op alle waarneembare variabelen die ruimtelijk gecorreleerd zijn.

Wanneer R en T vrij klein zijn, kan het voorkomen dat het aantal waarnemingen in verhouding tot het aantal waarneembare variabelen klein wordt. In een dergelijke situatie kan men veelal het eigenlijke aantal indicatoren vervangen door een aantal gecombineerde indicatoren van kleinere omvang.

Er zijn opgemerkt dat ruimtelijke auto- of cross-correlatie ten aanzien van de storingen verdisconteerd kan worden met de door Hordijk (1974) beschreven methoden. Deze kunnen ook worden gebruikt om met ruimtelijke auto-correlatie in de meetfouten van de exogene variabelen rekening te houden.

Laten we nu aandacht schenken aan het probleem van de temporele auto-correlatie, om te beginnen in het structurele model. De benadering die we hier zullen beschrijven is geïnspireerd door en vormt een uitbreiding van Jöreskog (1977). We beschouwen opnieuw model (15). We nemen aan dat de vector van storingen $u_{r,t}$ bestaat uit twee ongecorrleerde componenten: een vector van regionale elementen μ_r , welke constant zijn in de tijd en een vector van elementen $\omega_{r,t}$ welke zich gedragen volgens een eerste orde Markov schema:

$$u_{r,t} = \mu_r + \omega_{r,t} \quad (16)$$

met

$$\omega_{r,t} = D \omega_{r,t-1} + \zeta_{r,t} \quad (17)$$

met D een diagonale matrix:

$$D = \text{diag} (d_1, d_2, \dots, d_m), \quad |d_i| < 1, \quad i=1,2,\dots,m \quad (18)$$

De storingen $\zeta_{r,t}$ zijn nu ongecorrleerd zowel in als tussen de perioden. Eventuele correlatie binnen de perioden zou het gevolg zijn van ruimtelijke correlatie, waarvan wij veronderstellen, dat deze in het model verdisconteerd is. Eventuele correlatie tussen de perioden zou het gevolg zijn van ruimtelijke correlatie met vertraagde variabelen en/of van temporele autocorrelatie.

Met behulp van Kronecker produkten $\otimes$ kunnen vergelijkingen (15a) voor $t=1,2,\dots,T$ als één vergelijking geschreven worden:

$$(I \otimes B) \eta_r = (I \otimes \Gamma^1) \xi_r + \tilde{\Gamma}^2 \tilde{\xi}_r + u_r, \quad r=1,2,\dots,R \quad (19)$$

met:

$$\left. \begin{aligned} \eta_r &= (\eta_{r,1}^T, \eta_{r,2}^T, \dots, \eta_{r,T}^T)^T \\ \xi_r &= (\xi_{r,1}^T, \xi_{r,2}^T, \dots, \xi_{r,T}^T)^T \\ u_r &= (u_{r,1}^T, u_{r,2}^T, \dots, u_{r,T}^T)^T \\ \tilde{\Gamma}^2 &= ((\Gamma_1^2)^T, (\Gamma_2^2)^T, \dots, (\Gamma_T^2)^T)^T \end{aligned} \right\} \quad (20)$$

In (19) is I de $T \times T$ eenheidsmatrix. Gemakshalve laten we nu de index r ook weg.

Vergelijking (19) wordt nu:

$$(I \otimes B) \eta = (I \otimes \Gamma^1) \xi + \tilde{\Gamma}^2 \tilde{\xi} + u, \quad (21)$$

en vergelijking (17) :

$$\omega_t = D \omega_{t-1} + \zeta_t \quad (22)$$

We schrijven nu model (21) op de volgende wijze als een LISREL model:

$$\begin{bmatrix} I \otimes B & J \\ 0 & A \end{bmatrix} \begin{bmatrix} \eta \\ \omega \end{bmatrix} = \begin{bmatrix} I \otimes \Gamma^1 & \tilde{\Gamma}^2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \tilde{\xi} \end{bmatrix} + \begin{bmatrix} I \otimes \tilde{I} & 0 \\ 0 & I \otimes \tilde{I} \end{bmatrix} \begin{bmatrix} \mu \\ \zeta \end{bmatrix} \quad (23)$$

met:

μ ($m \times 1$) en ζ ($mT \times 1$) vectoren van storingen;

$\tilde{\xi}$ een kolomvector met T elementen gelijk aan één;

$\tilde{I}$ de $m \times m$ eenheidsmatrix

$\omega = (\omega_1^T, \omega_2^T, \dots, \omega_T^T)$ met ω_i^T een m dimensionale vector

A een $mT \times mT$ blokmatrix van de vorm

$$A = \begin{bmatrix} I & 0 & 0 & \dots & 0 & 0 \\ -D & I & 0 & \dots & 0 & 0 \\ 0 & -D & I & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & -D & I \end{bmatrix} \quad (24)$$

J een blokmatrix met $-D$ op de eerste onder-diagonaal en 0 matrices elders; dat wil zeggen J is de matrix A zonder de diagonale matrices I .

Het meetmodel voor de gecombineerde $(\eta^T \omega^T)^T$ vector wordt nu gespecificeerd als :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} \Lambda_y & 0 & \dots & 0 & \bar{\Lambda} & 0 & \dots & 0 \\ 0 & \Lambda_y & \dots & 0 & 0 & \bar{\Lambda} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Lambda_y & 0 & 0 & \dots & \bar{\Lambda} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_T \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_T \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_T \end{bmatrix}$$

waarbij $\bar{\Lambda}$ dezelfde structuur als Λ_y heeft met onbekende coëfficiënten $\bar{\lambda}$.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix} = \begin{bmatrix} \Lambda_x & 0 & \dots & 0 \\ 0 & \Lambda_x & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Lambda_x \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_T \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_T \end{bmatrix}$$

Tenslotte, het meetmodel voor de $\tilde{\xi}$ variabelen heeft de gebruikelijke vorm.

De correlaties tussen de ε 's en δ 's voor verschillende perioden kunnen tot uitdrukking worden gebracht in θ_ε en θ_δ .

Als het boven beschreven model niet geïdentificeerd is, kan men om te beginnen de matrices $\bar{\Lambda}$ en/of θ_ε en θ_δ beperken. Bijvoorbeeld θ_ε kan diagonaal gespecificeerd worden.

Uiteindelijk kan de methode helemaal niet bruikbaar zijn vanwege identificatieproblemen.

In een dergelijke situatie kan men tijdsspecifieke effecten verdisconteren met behulp van covariantie-analyse. Daartoe introduceert men dummy variabelen in de vector $\xi_{r,t}$ als volgt :

$$\xi_{r,t} = \begin{cases} 1 & \text{voor periode } t \\ 0 & \text{voor periode } s, \quad s \neq t \end{cases} \quad r=1,2,\dots,R; \quad t=1,2,\dots,T \quad (25)$$

Het gebruik van dummy variabelen heeft echter ernstige bezwaren (Maddala, 1971). Aan deze bezwaren kan men tegemoet komen door het gebruik van 'reële' tijdsspecifieke variabelen in plaats van dummy variabelen (zie Folmer, 1981).

Bovenstaande aanpak heeft zijn empirische merites bewezen in een studie naar de regionale effectiviteit van industrialisatiebeleid in Nederland (zie Folmer en Nijkamp, 1982).

5. AANVULLENDE OPMERKINGEN

We willen hier nog wijzen op een tweetal benaderingswijze van ruimtelijke auto- en cross-correlatie welke voor LISREL modellen interessante uitgangspunten bieden. In de eerste plaats, Hordijk en Nijkamp (1978) beschrijven het gebruik van Markov schema's voor ruimtelijke-temporele correlatie. Ten tweede, Streitberg (1979) beschrijft voor het standaard-lineaire model zonder latente variabelen de ontbinding van de covariantie-structuur als informatie beschikbaar is over een afhankelijkheidsstructuur van de regio's. Het voorgaande toont aan dat er op het terrein van LISREL-modellen met ruimtelijke auto- en cross-correlatie nog wel speurwerk mogelijk is.

References

- Cliff, A.D., J.K. Ord (1973): Spatial Autocorrelation. London: Pion.
- Droth, W., M.F. Fischer (1981): Zur Theoriebildung und Theorietestung. Eine Diskussion von Grundlagenproblemen am Beispiel der Sozialraum-analyse. In Ostheider, M., Steiner D. (eds.): Theorie und Quantitative Methodik in der Geographie, Zürcher Geographische Schriften. Heft 1. Geographisches Institut Eidgenössische Technische Hochschule Zürich.
- Dijkstra, T.K. (1981) : Latent Variables in Linear Stochastic Models. Ph. D. University of Groningen.
- Folmer, H. (1981) Measurement of the Effects of Regional Policy Instruments by means of Linear Structural Equation Models and Panel Data, Environment and Planning A 13, 1435-1448.
- Folmer, H., P. Nijkamp (1982): Linear Structural Equation Models with Spatio-Temporal Auto- and Cross-correlation, Research Memorandum, Economische Faculteit, Vrije Universiteit, Amsterdam.

- Folmer, H., J. Oosterhaven (1982) : Measurement of the Effects of Dutch Regional Socio-economic Policy. In Kuklinski, A., J.G. Lambooy (eds.): Regional Planning and Regional Policies, Experiences and Prospects: A Polish-Dutch Seminar. The Hague: Mouton. (forthcoming).
- Goldberger, A.S., O.P. Duncan (eds.) (1977) : Structural Equation Models in the Social Sciences. New York : Seminar Press.
- Hagget, P., A.D. Cliff, A. Frey (1977) : Locational Methods. London: Arnold.
- Hempel, C.G. (1958) The Theoreticians Dilemma, A Study in the Logic of Theory Construction. In : H. Feigl, M. Scriven, G. Maxwell, Philosophy of Science II, Minneapolis: University of Minnesota Press.
- Hordijk, L., P. Nijkamp (1977): Dynamic Models of Spatial Autocorrelation. Environment and Planning A 9, 505-519.
- Hordijk, L., P. Nijkamp (1978): Estimation of Spatio-temporal Models. In Karlqvist, A. and others (eds.): Spatial Interaction Theory and Planning Models, 177-200. Amsterdam : North-Holland.
- Insee, (1978): Annales de l'INSEE, issue 30-31 (April-September) (Proceedings of the Conference on Econometrics of Panel Data, August 22-24, 1977): Paris.
- Jöreskog, K.G. (1973a): A General Method for Estimating a Linear Structural Equation System. In: Goldberger, A.S., and O.P. Duncan (eds.): Structural Equation Models in the Social Sciences. New York: Seminar Press, 85-112.
- Jöreskog, K.G. (1973b): Analysis of Covariance Structures. In: Krishnaiah (ed.): Multivariate Analysis-III. New York: Academic Press, 263-285.
- Jöreskog, K.G. (1977): Structural Equation Models in the Social Sciences: Specification, Estimation and Testing. In: Krishnaiah (ed.): Application of Statistics, Amsterdam: North-Holland, 265-287.
- Jöreskog, K.G., D. Sörbom (1981): LISREL V. Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods. Mimeographed by Department of Statistics, University of Uppsala, Sweden.
- Leamer, E.E. (1978): Specification Searchers. Ad hoc Inference with Non-experimental Data. New York: Wiley.
- Maddala, G.S. (1971): The Use of Variance Component Models in Pooling Cross Section and Time Series Data. Econometrica 39, 341-358.
- Martin, R.L., J.E. Oeppen (1975). The Identification of Regional Forecasting Models Using Space-time Correlation Functions. Trans. Inst. Br. Geographers 66, 95-118.
- Silvey, S.D. (1970): Statistical Inference. Harmondsworth, Middlx: Penguin Books.
- Streitberg, B. (1979): Multivariate Models of Dependent Spatial Data. In Ketellapper, R.M., C.P. Bartels (eds.): Exploratory Analysis of Spatial Data, 139-178. Boston: Martinus Nijhoff, 139-179.

